EDGE-BASED METHOD FOR SHARP REGION EXTRACTION FROM LOW DEPTH OF FIELD IMAGES

Natalia Neverova and Hubert Konik

Laboratoire Hubert Curien – UMR CNRS 5516, University Jean Monnet, Saint-Étienne, France
Email: natalia.neverova@etu.univ-st-etienne.fr, hubert.konik@univ-st-etienne.fr

ABSTRACT

This paper presents a method for extracting blur/sharp regions of interest (ROI) that benefits of using a combination of edge and region based approaches. It can be considered as a preliminary step for many vision applications tending to focus only on the most salient areas in low depth-of-field images. To localize focused regions, we first classify each edge as either sharp or blurred based on gradient profile width estimation. Then a mean shift oversegmentation allows to label each region using the density of marked edge pixels inside. Finally, the proposed algorithm is tested on a dataset of high resolution images and the results are compared with the manually established ground truth. It is shown that the given method outperforms known state-of-the-art techniques in terms of F-measure. The robustness of the method is confirmed by means of additional experiments on images with different values of defocus degree.

Index Terms— Saliency, segmentation, edge detection, blur/sharp estimation

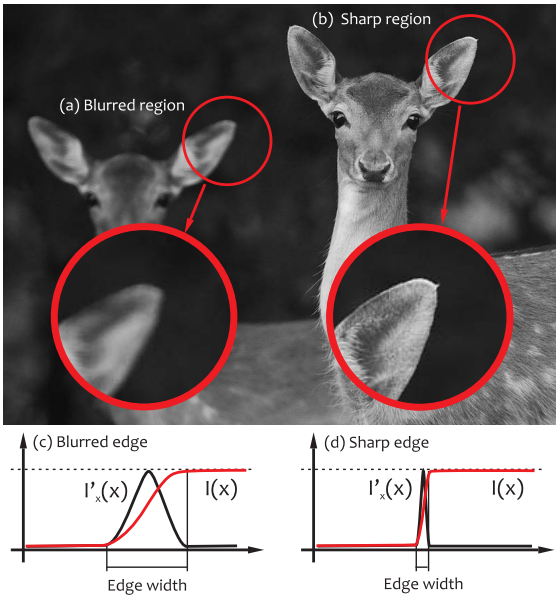

Fig. 1. (a) Blurred and (b) sharp regions with corresponding shapes of (c) blurred and (d) sharp edge intensity profiles.

1. INTRODUCTION

Visual attention aspects play an important role in many computer vision applications. The initial objective of saliency estimation is to locate regions of interest (ROI) in images in order to increase efficiency of content-based image retrieval, compression, auto-cropping or object recognition. The idea is based on a commonly used assumption that the most amount of information about an image content is concentrated in the most salient areas of the image. Consequently, a large number of methods in the field of visual communications and image processing could be improved if the location of such specific regions is known a priori. Generally speaking, the most classical models define saliency based on color, texture and orientation. However, aside from the fact that the generated saliency maps usually have a lower resolution compared with input images, these models do not take into account some important visual effects. For instance, in the case of so-called low depth-of-field (DOF) images, objects in focus appear to be sharp and visually attractive, while blurred background does not immediately draw the conscious attention of the viewer.

In that context, [1] have recently experimentally shown this precise influence of the sharp/blur aspect of an image part on its saliency. These results indicate that blur information might be integrated in attention models to efficiently improve the extraction of salient regions. In fact, sharp objects tend to capture attention irrespective of intensity, color or contrast.

There have been a number of methods proposed for salient region extraction, some of which are in the context of low depth of field images. All of them, to varying degrees, exploit an assumption that focused regions contain more high-frequency components than the defocused background. For example, [2] propose a computational Markov random field formulation for generating a defocus map, where each image pixel is then considered independently and claimed to be either salient or not based on a preliminary set value of threshold. Unfortunately, there is no detailed information on

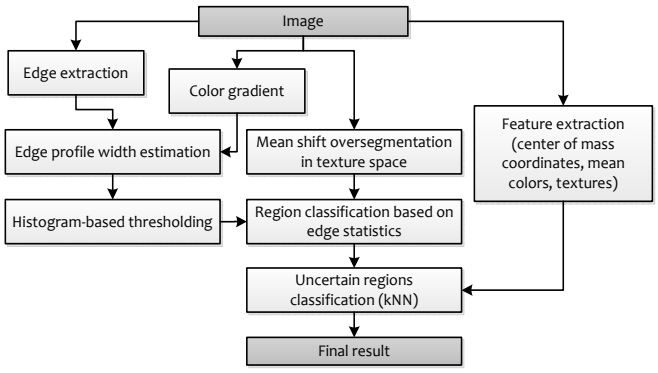

Fig. 2. Block diagram of our algorithm.

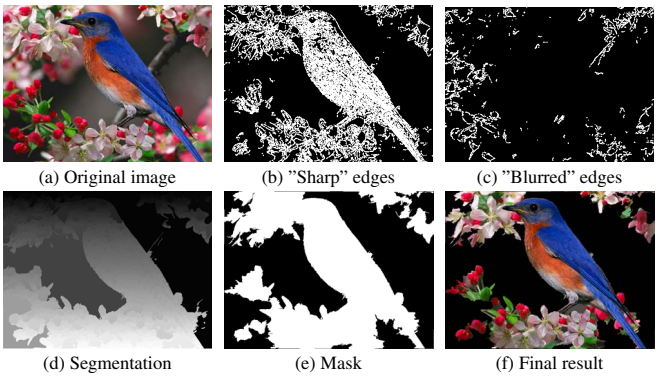

Fig. 3. Example of intermediate results.

the method performance or comparison with alternative techniques. In [3], the authors introduce the orientation angle of the gradient map to model the focus energy and evaluate the boundary and region saliencies respectively. Finally, some post-treatments are proposed to isolate sharp objects using a boundary linking method (at cost of processing time). In [4], the authors present a machine learning based approach to identify focused objects using multiple segmentations and visual descriptors computation. However, all these methods do not explicitly involve edge processing, while, according to [5], the edge information may appear to be especially important for the precise saliency-based segmentation.

In this paper, we present a simple and effective bottom-up precise segmentation of low depth-of-field images. First, we use the shape of the edge gradient profile at each particular edge point to classify edge sharpness. After that, an image mean shift oversegmentation in texture space coupled with a set of classification rules allows to precisely localize the focused objects on the blurry background.

The remainder of this paper is organized as follows: section 2 describes the proposed framework. Then, section 3 provides some experimental results and illustrates the method performance in comparison with several alternative state-of-the-art techniques over an established ground truth of images. Finally, conclusions and future work are provided in section 4.

2. OUR METHOD

The proposed approach is based on the idea that the edge width (and the corresponding shape of the edge gradient profile) at each particular edge point can be considered as a function of the degree of blur in a given image region. In our case, by the term "gradient profile" we mean a distribution of image gradient magnitude in the vicinity of a local maximum ("edge pixel") taken in the direction of the maximum rate of change at this point.

According to [6], the shape of the gradient magnitude distribution around edges in natural images is reasonably stable and generally obeys the generalized Gaussian distribution with parameter $\lambda \approx 1.6$ regardless of the image resolution. The sharper the edge, the smaller is the variance of the edge gradient profile distribution (Fig. 1).

In this work, we estimate the edge sharpness at each edge point (x_0, y_0) calculating the standard deviation of the edge gradient profile $p(x_0, y_0)$ [6].

$$\sigma(p(x_0,y_0)) = \sqrt{\frac{\sum_{(x,y)\in p(x_0,y_0)} ||\mathbf{G}(x,y)|| \cdot l^2(x,y)}{\sum_{(x,y)\in p(x_0,y_0)} ||\mathbf{G}(x,y)||}}, \quad (1)$$

where $||\mathbf{G}(x,y)||$ is the gradient magnitude and $l(x,y)$ is the distance between each point (x,y) and the edge point.

An overview of our method is sketched in Fig. 2. Some intermediate results obtained in each step of the process for one image from our database are provided in Fig. 3. All steps will be discussed in more detail in the following sections.

2.1. Edge detection

Since the quality of the edge extraction and localization is crucial for our method, we have adapted the Canny detector to be capable of using color information.

The Canny detector requires computing the convolution of an image with a Gaussian K_{σ_0}, then with its first derivative along each spatial direction $\{x, y\}$ [7]. It can easily be done for each color plane in the straightforward way, using RMS to combine the obtained results for three color components. Let us denote the result of double convolution as $F_{c,d}$, where c stands for the color component and d for the spatial direction:

$$F_{\{r,g,b\},x} = [(\{R,G,B\} * K_{\sigma_0}(x)) * K_{\sigma_0}(y)] * \frac{\partial}{\partial x} K_{\sigma_0}(x);$$

$$F_{\{r,g,b\},y} = [(\{R,G,B\} * K_{\sigma_0}(x)) * K_{\sigma_0}(y)] * \frac{\partial}{\partial y} K_{\sigma_0}(y).$$

Letting $\mathbf{r}$, $\mathbf{g}$ and $\mathbf{b}$ be unit vectors along the R, G and B axes of RGB color space, we define the following quantities [8]:

$$\mathbf{f}_{\{\mathbf{x},\mathbf{y}\}} = F_{r,\{x,y\}}\mathbf{r} + F_{g,\{x,y\}}\mathbf{g} + F_{b,\{x,y\}}\mathbf{b},$$

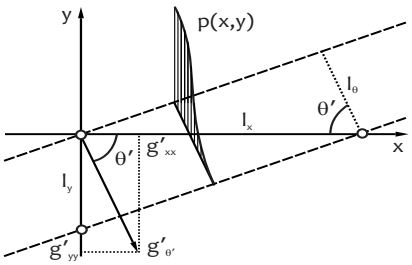

Fig. 4. Edge gradient calculation.

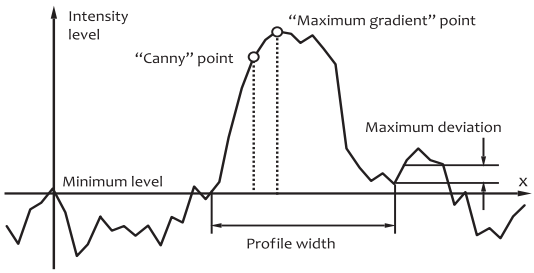

Fig. 5. Width estimation of the edge gradient profile.

$$g_{\{xx,yy\}} = ||\mathbf{f}_{\{\mathbf{x},\mathbf{y}\}}|| = \sqrt{F^2_{r,\{x,y\}} + F^2_{g,\{x,y\}} + F^2_{b,\{x,y\}}},$$

$$g_{xy} = \sqrt{\mathbf{f_x} \cdot \mathbf{f_y}} = \sqrt{F_{r,x}F_{r,y} + F_{g,x}F_{g,y} + F_{b,x}F_{b,y}}.$$

In this case, considering the result of convolution to be approximately corresponding to the image gradient, the direction of maximum rate of change is calculated by [8]

$$\theta(x,y) = \frac{1}{2}\arctan\left(\frac{2g^2_{xy}}{g^2_{xx} - g^2_{yy}}\right), \tag{2}$$

The magnitude of the rate of change ("edge strength") in the given direction can be calculated by

$$g_\theta = \sqrt{\frac{(g^2_{xx}+ g^2_{yy}) + (g^2_{xx} - g^2_{yy})\cos 2\theta + 2g^2_{xy}\sin 2\theta}{2}}. \tag{3}$$

The final steps of edge extraction are performed in standard way following the procedure of the original Canny algorithm.

2.2. Gradient profile analysis

To estimate the edge width at each edge point, it is necessary to analyze an undistorted gradient distribution, where the original shape of the gradient profile around edges is preserved (consequently, the smooth gradient map obtained in the previous step cannot be used). Here we calculate a simple image gradient and denote it as $g'_{\theta'}$ [8].

In the next step the proposed algorithm requires sampling the gradient magnitude $g'_{\theta'}$ along the direction of maximum rate of change θ'. To avoid additional computations, [9] proposed to analyze the gradient magnitude profile in the horizontal and vertical directions and then, if needed, use their properties on the specified maximum gradient direction. Given the gradient distributions along the horizontal and vertical axis (g'_{xx} and g'_{yy} respectively) and edge width in the corresponding directions (l_x and l_y respectively), the actual edge width $l_{\theta'}$ can be calculated by (see Fig. 4)

$$l_{\theta'} = \frac{g'_{xx}}{g'_{\theta'}} \cdot l_x = \frac{g'_{yy}}{g'_{\theta'}} \cdot l_y. \tag{4}$$

The same relation holds for the standard deviation $\sigma_{\theta'}$ representing edge "sharpness" (the standard deviation in this case is defined by (1)):

$$\sigma_{\theta'} = \frac{g'_{xx}}{g'_{\theta'}} \cdot \sigma_x(p_x(x_0)) = \frac{g'_{yy}}{g'_{\theta'}} \cdot \sigma_y(p_y(y_0)), \tag{5}$$

where $\sigma_x(p_x(x_0))$ and $\sigma_y(p_y(y_0))$ correspond to the standard deviation of gradient profiles along the horizontal and vertical directions respectively, which are calculated by (see [9])

$$\sigma_x(p_x(x_0)) = \sqrt{(\sum_{x\in p(x_0)} g'_{xx}l^2_x)/(\sum_{x\in p(x_0)} g'_{xx})},$$

$$\sigma_y(p_y(y_0)) = \sqrt{(\sum_{y\in p(y_0)} g'_{yy}l^2_y)/(\sum_{y\in p(y_0)} g'_{yy})}.$$

Since for our purposes we localize edges using the Canny detector, obtained edge points do not necessarily correspond to pixels with maximum values of the gradient magnitude (Fig. 5). For this reason, before performing the gradient profile analysis, we find a local maximum in the chosen gradient direction. If this maximum point cannot be reached within the specified distance from a "canny" pixel, the latter one is taken out of consideration as an edge detector error.

As soon as the local maximum is reached, the gradient profile can be analyzed by exploring magnitude values on both side of this point along the given direction. Fig. 5 illustrates the idea how to do it using two stopping criteria.

First, we consider only pixels with intensities not smaller than the selected minimum level (as a fraction of the maximum value). Second, it is necessary to "cut" the profile if there is another edge nearby setting a maximum positive deviation threshold. In our implementation we set these two parameters equal to 10% and 0.05 respectively.

Due to presence of noise, influence of nearby edges, errors of the Canny detector, etc., the $\sigma_{\theta'}$ values even for neighboring edge pixels can vary greatly. To get a more reliable classification and compensate for these negative effects, for further processing we replace the actual values of $\sigma_{\theta'}$ with an average value along each edge.

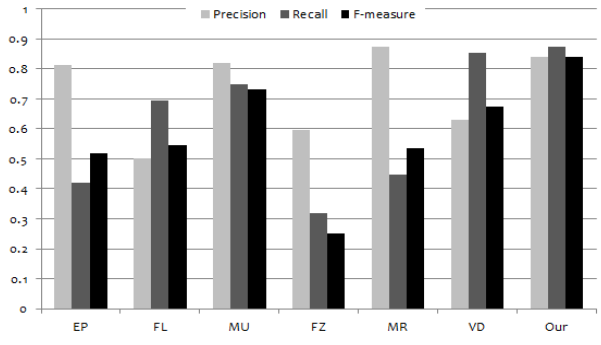

Fig. 6. Comparison of the results obtained with different methods on the given dataset

2.3. Segmentation

Since after the segmentation step it will be necessary to assign a label to each region (either "sharp" or "blurred") considering it as a single entity, the undersegmentation and combining different objects into one region are unacceptable. To avoid this problem, we oversegment the image using a mean shift algorithm [10], which provides accurate boundaries between regions. The segmentation is performed in texture space, where each feature plane is obtained with one of Laws' texture energy filters applied to each of RGB color planes.

2.4. Classification

In the final stage of the algorithm we classify all previously obtained regions into 2 classes ("sharp" or "blurred") based on the presence of "sharp" and "blurred" edges inside each segment. For each region we calculate a set of the following features: center of mass coordinates, region area, average RGB values, average values for each Laws' textures component for intensity channel.
The further classification is based on the following logic:

- If there is only one kind of edge pixels inside the region, the region area is greater than 64 pixels and the density of edge pixels is greater than the minimum value between 0.015 and 90% of the average edge pixel density, this region is labeled in accordance with edge labels.
- If one kind of edge pixels is in absolute majority (the ratio between the number of pixels belonging to two categories is greater than 10), the region area is greater than 64 pixels and the edge pixel density is greater than the maximum value between 0.015 and 90% of the average edge pixel density, region label will correspond to the label of the majority of edge pixels.
- Else (small or sparse regions, containing different kinds of edge pixels), the label is assigned based on labels of k nearest neighbors already having labels (the classification is based on the set of features defined above).

3. EXPERIMENTAL RESULTS

The proposed algorithm has been tested on a dataset of 112 high resolution images having a great variety in content, colors, textures, numbers of sharp and blurred regions, their area and location[1]. All parameters were optimized in advance and the experiments were conducted automatically, with no manual tuning from image to image. To illustrate the performance of the proposed algorithm in comparison with known state-of-the-art techniques, we used the same dataset to test 6 alternative methods: Achanta et al. [11], Achanta and Susstrunk [12], Graf et al. [13], Zhang et al. [14], Kim [15], Li and Ngan [16]. We refer these methods as EP, FL, MU, FZ, MR an VD, respectively.
The performance of all algorithms was estimated in terms of recall, precision and F-measure (with $\beta = 1$). Fig. 6 contains the summarized diagram, which demonstrates that our algorithm outperforms the alternative techniques mentioned above on the given dataset and in terms of the given quality measures. The main advantage of our method is a possibility to reach great values of recall and precision simultaneously and in the absolute majority of cases. Examples of images from the dataset together with the manually created ground truth binary masks and masks obtained with our and state-of-the-art methods are shown in Fig. 8. Subjectively, our approach yields the most visually plausible results regardless of image size, content and complexity.
To evaluate the robustness of our method, we have conducted some additional experiments on images with different values of defocus degree. We used several image sequences with manually blurred background (using Gaussian filters with different parameters), as well as images of several scenes taken with a camera with differing aperture settings. Defocus degree for each image of each sequence was calculated by

$$D = \frac{\text{mean}_{sharp}(g'_{\theta'})}{\text{mean}_{blurred}(g'_{\theta'})}.$$

A typical graph of F-measure vs. defocus degree for one image sequence is shown in Fig. 7. It illustrates the fact that starting from a certain relative degree of blur between "sharp" and "blurred" objects present within the scene, our method provides confident segmentation with an asymptotically increasing F-measure. The observed general behavior holds for all tested image sequencies, either synthetic or natural.

4. CONCLUSION

In this paper, we presented an efficient unsupervised method for salient object extraction in low depth-of-field images. Our approach is based on edge analysis and the assumption that blurred edges are generally wider than the sharp ones. It was experimentally shown that in comparison with alternative

[1] http://dossier.univ-st-etienne.fr/konikhub/public/Database/blur-sharp.zip

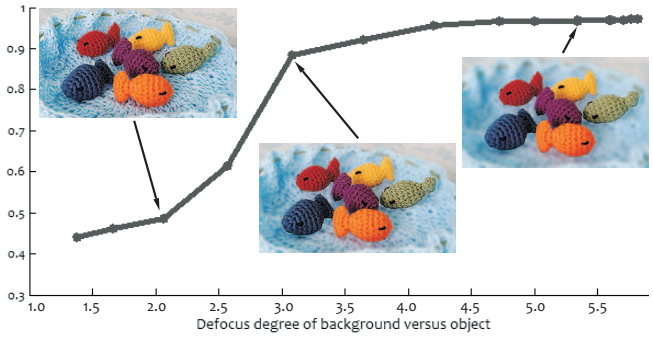

Fig. 7. Illustration of our algorithm performance on a sequence of images with different defocus degrees.

frameworks, this method achieves higher recall and precision rates simultaneously and in the absolute majority of cases.
In future work, we first aim to explore the effect of restricting some image processing techniques to selected image regions, which is necessary in such applications as saliency-aware image and video coding and retrieval. Moreover, encouraged by the high performance of the proposed method, we plan to introduce its results to improve visual attention models with a general framework taking this blur/sharp aspect into account.

5. REFERENCES

[1] Hubert Konik, Rizwan Ahmed Khan, and Eric Dinet, "Visual attention: effects of blur," *IEEE International Conference On Image Processing (ICIP)*, pp. 3350–3353, 2011.

[2] Yu-Wing Tai and Michael S. Brown, "Single image defocus map estimation using local contrast prior," *IEEE International Conference On Image Processing (ICIP)*, pp. 1797–1800, 2009.

[3] Zhi Liu, Weiwei Li, Liquan Shen, Zhongmin Han, and Zhaoyang Zhang, "Automatic segmentation of focused objects from images with low depth of field," *Pattern Recognition Letters*, vol. 31, no. 7, pp. 572–581, 2010.

[4] Hongliang Li and King N. Ngan, "Learning to Extract Focused Objects from Low DOF Images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 11, pp. 1571–1580, 2011.

[5] Cairong Zhao, ChuanCai Liu, Zhihui Lai, and Jingyu Yang, "Sparse Embedding Visual Attention Systems Combined with Edge Information," *IEEE International Conference on Pattern Recognition (ICPR)*, pp. 3432–3435, 2010.

[6] Jian Sun, Zongben Xu, and Heung-Yeung Shum, "Image super-resolution using gradient profile prior," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.

[7] John Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[8] Silvano Di Zenzo, "A note on the gradient of a multi-image," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 1, pp. 116–125, May 1986.

[9] Luhong Liang, Jianhua Chen, Siwei Ma, Debin Zhao, and Wen Gao, "A no-reference perceptual blur metric using histogram of gradient profile sharpness," *IEEE International Conference On Image Processing (ICIP)*, pp. 4369–4372, 2009.

[10] Christopher M. Christoudias, Bogdan Georgescu, and Peter Meer, "Synergism in low level vision," *IEEE International Conference on Pattern Recognition (ICPR)*, pp. 150–155, 2002.

[11] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Süsstrunk, "Frequency-tuned salient region detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1597–1604, 2009.

[12] Radhakrishna Achanta and Sabine Süsstrunk, "Saliency detection using maximum symmetric surround," *IEEE International Conference On Image Processing (ICIP)*, pp. 2653–2656, 2010.

[13] Franz Graf, Hans-Peter Kriegel, and Michael Weiler, "Robust segmentation of relevant regions in low depth of field images," *IEEE International Conference on Image Processing (ICIP)*, pp. 2861–2864, 2011.

[14] KeDai Zhang, HanQing Lu, ZhenYu Wang, Qi Zhao, and MiYi Duan, "A fuzzy segmentation of salient region of interest in low depth of field image," *Proceedings of the 13th international conference on Multimedia Modeling - Volume Part I*, pp. 782–791, 2006.

[15] Changick Kim, "Segmenting a low-depth-of-field image using morphological filters and region merging," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1503–1511, Oct. 2005.

[16] Hongliang Li and King N. Ngan, "Unsupervized video segmentation with low depth of field," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1742–1751, 2007.

Fig. 8. Examples of the results given by different methods together with the established ground truth.